

# THE PSYCHOLOGY OF CONFIDENCE— An Experimental Inquiry

BY  
WM. CLARK TROW, Ph. D.

ARCHIVES OF PSYCHOLOGY

Edited by  
R. S. WOODWORTH

No. 67. Nov. 1923

PS: G. E. STECHERT & CO.: London (2 Star Yard, Carey St., W. C.);  
Paris (16 rue de Condé).

BF  
21  
A7  
no. 67



Northeastern University  
Math/Psychology  
Library

4

# THE PSYCHOLOGY OF CONFIDENCE— An Experimental Inquiry

BY  
WM. CLARK TROW, Ph.D.

ARCHIVES OF PSYCHOLOGY

Edited by  
R. S. WOODWORTH

No. 67. Nov. 1923

AGENTS: G. E. STECHERT & CO.: London (2 Star Yard, Carey St., W. C.);  
Paris (16 rue de Condé).

Psych  
BF  
21  
A.11  
no. 69

## CONTENTS

	Page
Chapter I. Introduction .....	3
1. Historical Background.	
2. Setting of the Problem.	
Chapter II. Procedure .....	7
1. General.	
2. The Scale of Confidence.	
3. The Rating of Confidence.	
4. Self-estimates and Those of Others.	
5. Detailed Procedure.	
6. The Method of Scoring Confidence.	
Chapter III. Confidence and Achievement, Their Distribution and Correlation .....	26
Chapter IV. Other Relationships.....	29
1. Confidence Scores and Ratings.	
2. Confidence and Achievement Scores and Intelligence.	
3. Comparative Value of the Separate Measures.	
Chapter V. Correctness and the Degrees of Confidence .....	33
Chapter VI. Confidence as a Character Trait.....	37
Chapter VII. Conclusions .....	40
Appendix:	
A. The Indicators Used.....	42
B. Bibliography .....	46

# The Psychology of Confidence— An Experimental Inquiry<sup>1</sup>

## I.

### INTRODUCTION

#### 1. HISTORICAL BACKGROUND

6 Feb. 1957.

The widespread commercialization of the word psychology indicates, among other things, a more or less scientific interest on the part of people generally in the activity of the human organism as a whole. If fact is to be gradually substituted for fable, it is the task of experimental psychology to supply the facts, baffling though they may be; and no *terra incognita* would seem to afford more opportunity for such inquiry than that of character traits. It is with a small portion of this field, the question of confidence, that the present study is concerned; and while it will, no doubt, eventually be possible to follow it through its various business, industrial, educational, abnormal and social applications, no attempt is made to do this here.

Previous inquiry on the subject is reviewed briefly under the following heads: (1) Non-experimental; (2) Introspective; (3) Quantitative. If there be those who take exception to the classification, they are at liberty to alter it; the classification is not the important thing.

#### (1) NON-EXPERIMENTAL INQUIRY

The earlier psychological writing on the subject of belief, confidence, or assurance, as it is variously termed, has been ably reviewed by Lindsay<sup>2</sup> and by Okabe, and need only be mentioned here.

Bain calls belief a mental state which, though involving the intellect and the feelings, is essentially related to activity in that what we

<sup>1</sup>The writer is greatly indebted to the members of the Department of Psychology of Columbia University for their many helpful suggestions and criticisms and to the students who gave their time as subjects for this experimentation.

<sup>2</sup>The Bibliography in the Appendix includes the references for all writings cited in this chapter.

believe we act upon. Brentano makes belief a separate, unanalyzable mental element. Bagehot speaks of the emotion of conviction, acquiescence or consent, while James makes belief a kind of "feeling more allied to the emotions than anything else . . . a psychic attitude toward a proposition." He points out that the opposite of belief is doubt, not disbelief. Stout calls it the "yes-no consciousness," distinguishes it from simple apprehension, sees its relation to desire and recognizes all manner of gradations proportioned to the difficulty of substituting for a thought its alternative. Sully also recognizes degrees of doubt and belief, as well as differences of individual "temperament"—the "energetic" and the "cautious."

Possibly the behavioristic position<sup>1</sup> should be mentioned here, for it is as yet speculative rather than scientific, making thinking the action of language mechanisms, judgment or decision the dying away of intraorganic stimuli,<sup>2</sup> and belief a "positive reaction toward."

## (2) INTROSPECTIVE INQUIRY

Roback presented diverse statements from many authors to seven subjects from whose replies he concluded that belief or disbelief is conditioned rather by "the congruity of the imagery induced by the passages with the memory images of a similar situation actually experienced," than by any logical aspect involved. The bodily feelings accompanying acceptance and rejection are also described. This is done likewise by Okabe, who uses the term "belief-disbelief consciousness." McDougall has made belief one of the "derived emotions," relating it, like Shand, to desire. Titchener connects it with the "feeling of reality" and quotes other writers who use this term. Cases in which the absence of confidence is the rule are familiar to those with clinical experience as, for example, the cases cited by Janet and classed as feelings of difficulty, of incapacity, of indecision, of irresolution, etc.

## (3) MORE OBJECTIVE, QUANTITATIVE INQUIRY

Turning from efforts to describe the belief consciousness, let us see what the results of various kinds of performance experimentation

---

<sup>1</sup>Upon being asked by letter what the behavioristic position is on the subject, Dr. Watson replied as follows: "I am afraid you have come to the wrong market on the subject of judgment, confidence, etc. They are not terms that would ordinarily be used by the behavior school at all." In view of this, it may be unfair to appeal to the camp of the enemy; however, Roback, in his "Behaviorism and Psychology" is suggestive on this subject.

<sup>2</sup>Watson, J. B. Notes of a lecture (unpublished) delivered at Teachers College, Columbia, 1923.

have been. Fullerton and Cattell in their psychophysical experimentation found, as early as 1892, that "some observers are not confident unless they are, in fact, right; while others are often confident when they are wrong." Griffing, in 1895, using the A, B, C, D scale in experimenting with the sensation and perception of dermal stimuli concluded, incidentally, that "the degree of confidence in the perception of intensive difference varies greatly for individuals," observers, when confident, ranging from one-third to one-fiftieth wrong, and that correctness is an independent variable. Henmon, having his subjects judge the length of lines, concluded in 1911 that the relation of confidence to accuracy seemed to be an individual matter without any well defined central tendency. Strong, using six series of advertisements for testing recognitive memory, concluded that some subjects had a "conservative temperament" or "do not like to take chances." The "conservative" individual makes practically no mistakes in his first choices, while the "optimistic" makes many. Metcalf, by the measure of speed, pressure, etc., in drawing figures, discovered that "certainty is usually found to go faster and with an accelerated rate of drawing, and with greater pressure." Most complete and satisfactory of all has been the work of Hollingworth, a part of whose experimentation is referred to in Chapter V.

## 2. SETTING OF THE PROBLEM

It will be seen from the above resumé, and from the other writings noted in the bibliography, how the point of view has gradually shifted, and how the trend of inquiry is aiming more or less carefully at the following points, which, it seems to me, should be clearly distinguished:

1. The subjective feeling of confidence, introspectively reported.
2. Its relation to desire, *i. e.*, confidence that the future will bring the fulfillment of one's desires,—optimism; or confidence in a cause or ideal,—faith; or confidence in some person or leader, etc.
3. Confidence in the correctness of one's judgments. This is the question to which the present experimentation is committed.
4. Confidence in one's self,—willingness to act overtly, in the social situation, on the basis of what confidence one has. Such action is necessarily one of the criteria for the judging of the

confidence of others; as may readily be seen, it is quite likely to be a false one.

5. Motor impulsiveness, speed of reaction or decision. This is pointed out by Filter in a most suggestive paper and is another of the untrustworthy criteria for judging the confidence of others.

Since there are such different items loosely classed under the term confidence, we are constantly confronted by the danger of aiming at the bear generally, and then standing off to see what the results are. Such valuable exploratory work as that of Dr. June Downey is a case in point. Certain assumptions are made, the validity of which is open to experimental investigation; for example, it might well be asked if one test situation is sufficient to brand an individual as possessed or not possessed of a certain trait or combination of traits.

Moore and Gilliland, in a similarly ingenious type of experiment endeavor to deal with something they call aggressiveness, but are handicapped by the inclusiveness of their definition, which they make "synonymous with personal force, initiative, assurance. It is understood as standing for that trait which, in combination with intelligence and reliability, goes far toward completing the essential personal requisites for success."

The difficulty is that personal force, initiative and assurance are very different things, possibly independent variables. We shall even find considerable difficulty in treating assurance by itself as a unit, to say nothing of the others. Furthermore, if an analogy is sought in intelligence, it should be remembered that the factors going to make up intelligence as represented, say, on a standard test, have long been an object of careful laboratory study, which is, as yet, almost entirely lacking in the case of the so-called character traits.

Perhaps the greatest difficulty in reducing character traits to a standardized test procedure is the influence of the social factor. As Hollingworth says,<sup>1</sup> these traits, like "cooperativeness and cheerfulness are functions of the circumstances in which a person is placed." Link,<sup>2</sup> whom he quotes, also stresses this same point.

The present study aims directly at one separate phase of the subject, *i. e.*, the confidence an individual may have in his own judgments. The method used, that of ascribing a degree of a confidence to a judgment, while presenting obvious difficulties, has nevertheless

<sup>1</sup>Judging Human Character, p. 146.

<sup>2</sup>Link, H. C.—Employment Psychology, p. 202, ff.



been used to advantage before in certain circumscribed types of experimentation. The method is here applied to diverse kinds of judgment situations, and, conducted with a larger number of subjects than heretofore. Hence it is possible to draw conclusions as to the influence of the type of situation which the subject judges, as well as to see more clearly the differences between the individuals themselves.

## II.

### PROCEDURE

#### 1. GENERAL

The sixteen indicators, which we shall call tests for convenience, realizing that they are not that technically speaking, were administered to forty-two subjects. Since, roughly speaking, each test consisted of twenty judgments upon which each subject's confidence was obtained, approximately 320 judgments were obtained from each subject and 13,440 from them all.

The tests were given to three rather distinct groups, so that ratings by each subject of those with whom he was acquainted might be obtained. Group I. consisted of fifteen male subjects, students in the experimental psychology course for undergraduates given in Columbia. Group II. consisted of fifteen male subjects, and Group III. of twelve female subjects, the latter two groups being graduate students in psychology. All subjects had had a minimum of a year's work in psychology; most had had a great deal more; one was a holder of the degree of doctor of philosophy. This psychological training was desirable in consideration of the type of experimentation.

The experiments were conducted during the spring of 1923 in the laboratory of Columbia University.

The tests will be discussed in the order in which they were given, though for greater ease in handling them they were tabulated in a slightly different order.

Nearly all directions were written, either being typed on 5 x 11 cards, as for Tests I. through VI., or appearing at the head of the test sheets, as for most of the remaining tests.

The first card each subject was shown was the following: During the experimentation which follows, most of the directions will be in written form for the sake of standardization and clarity. If you do not understand them at any point, do not hesitate to ask the

experimenter to clear up any obscurity, for such questions, as well as the time spent in reading the instructions, are no part of the experimentation proper. The aims of the experiments are various, but through them all runs one rather difficult requirement, namely, that you evaluate as carefully as possible your degree of confidence in the various situations. Four degrees of confidence are described on the next card. You may refer to that card as frequently as you care to, with a view to keeping the four degrees of confidence as constant as possible during the experimentation. Work at your normal rate of speed. In only one test is a speed record sought, and you will be told which one that is.

## 2. THE SCALE OF CONFIDENCE

A—*Confidence* means that you are *perfectly confident, absolutely certain*, as certain as you are that two plus two equals four, that you speak English, that you are reading these directions. It is the kind of confidence that admits of no thought of error, even against a contrary view of others.

B—*Confidence* means that you are *fairly confident, reasonably sure* of your judgment. You would be willing to bet on it (if you do bet), but would by no means wager all you have. If you should put up a reasonable sum and lose, you would probably say, "Well, I know I took a chance, but I didn't think I'd lose."

C—*Confidence* means that your judgment is made *with little confidence*; you are only *slightly certain*. You rather think so, though you would accept a contrary view, for you think such a contrary view *might* be superior to yours. If you had been willing to bet and had lost, you would probably have said, "Well, I didn't really have much idea I'd win."

D—*Confidence* means that your response is a *mere guess*. It is what might be called a fifty-fifty proposition, as for instance, that the sun will be covered by a cloud at noon Sunday, or that there are an even number of people in New York City. It is at the other extreme from A—Confidence. You would be perfectly willing to reverse your opinion, and then you would be as uncertain as you were before.

It is clearly recognized that such a scale presents certain dangers, yet for the purpose of this experimentation it has advantages which no other method has.

The order-of-merit arrangement used by Sumner,<sup>1</sup> though it offers a finer measure, would be unsatisfactory here since it does not give a record of absolute, only relative confidence. For example, two persons might make an almost identical rating of beliefs, and yet one might be highly confident of all of them, while the other might be exceedingly doubtful even of the ones at the top of the list. Furthermore, it would seem that comparatively few judgments in life are of this kind. Rather, things come one at a time. The wide range of materials, moreover, in this experimentation, do not permit of the use of this type of scale.

The method of placement on a scale, graphically, has certain advantages, too. It is novel; it overcomes the relative-confidence objection above, and has other merits.<sup>2</sup>

However, it was not used here, for such refinement in such diverse materials is almost impossible for the subject, and most judgments in life situations are not of this kind. Furthermore, though stimuli may vary gradually through many degrees, the evidence is not conclusive that confidence does likewise. It may, but this is an assumption.<sup>3</sup>

The method employed also makes assumptions, and has certain disadvantages, for the material, as is well recognized, does not lend itself readily to exact quantitative measurement. But these seem of

<sup>1</sup>Sumner, F. B.—A Statistical Study of Belief. Psychol. Rev., 1898, 5, 616-631.

<sup>2</sup>Hollingworth, H. L.—Judging Human Character, p. 105.

<sup>3</sup>E. K. Strong, in his "Introductory Psychology for Teachers, (Warwick & York, 1920) page 11, seemingly basing his conclusions in part on the work of Sumner, above referred to, draws up a tentative scale of belief, as follows:

99  $2+2=4$ .

73 There exists an all-wise Creator of the World.

47 A housefly has six feet.

21 The most honest man I know will be honest 10 years from now.

— 2 Blessed are the meek for they shall inherit the earth.

—22 Magna Charta was signed in 1512.

—53 It never rains but it pours.

—74 Only the good die young.

—99  $2+4=7$ .

Dr. Strong says, "If one wishes to determine, for example, how strongly he believes that 'dark-haired girls are prettier than light-haired ones,' he can compare it with those statements above, and so get a rating for it."

But suppose he is a scientist and *knows* that a housefly has six feet; and suppose he is likewise an atheist, as he might also be, and a cynic, as he might also conceivably be. What has become of the positive end of his scale of beliefs to "determine how strongly he believes" in the pulchritudinous superiority of the brunettes? This criticism is not due to the tentative character of the scale, that it is based on few cases, and the like, but to its inherent nature. Propositions cannot be used to measure the beliefs of an individual unless each constructs his own scale, for the confidence of different individuals in the same proposition varies from 100 to 0; and if disbelief is measured below on to -100.

less grave nature, and the advantages of sufficient weight to justify the method employed.

In the first place, it is a crude measure that has but four degrees on it. However, it seems doubtful if isolated judgments are capable of much closer refinement. The writer finds difficulty in locating finer distances with any feeling of satisfaction; and rarely, during the experimentation did a subject seem to feel any need of intermediate points on the scale.

In the second place, it assumes that the points are the same for all individuals. So far as the A- and D-judgments are concerned, there is probably no danger in this. With the B- and C-judgments there is possibly a little variation, but not so much as there would be with a greater number of degrees, probably.

On the positive side, we have several advantages. The scheme is workable in the wide diversity of situations of the experiment, and is readily grasped and employed by the subjects. It escapes most of the disadvantages of the other methods, and lends itself to statistical treatment.

### 3. THE RATING OF CONFIDENCE

#### TEST XVI.

This part of the experimental procedure was run through first so that the judgments made would not be on the basis of the experimentation. It was tabulated last so that it would group more easily with the tests for which there was no objective reference for the correctness of the responses.

#### *Materials:*

A set of cards  $2\frac{1}{2} \times 5\frac{1}{2}$  inches in size, each with the last name of a subject typed on it. Each member of each of the three groups rated the members of his group, including himself, for the quality of self-confidence, which was defined on the directions card, which read as follows:

#### *Directions:*

In rating the persons whose names appear on these cards, place the most self-confident at the left, the next most self-confident just to the right of it, and so on to the person who you think has the least confidence, whose card you will place at the extreme right. These judgments are confidential, so you need have no fear of making the arrangement exactly as you think it should be. Be as sure as possible that confidence is the trait you are rating, and not any other such as intelligence, humor, co-operativeness, scholarship, etc. To

assist in getting a uniformity of meaning for self-confidence the following suggestive definitions are included,—State of mind characterized by reliance on one's self, or one's circumstances, assurance. Confidence in the correctness of one's ideas or acts. Extent of adherence to one's opinions and beliefs—self-sufficiency in situations generally and willingness to take the lead.

The word "self-confidence" was used because it distinguished the trait in question from the other meaning of confident, namely, trustful and confiding, and also because it seemed the nearest thing to what the tests were after, a measurement of the confidence of the subjects in their own judgments and opinions.

When the arrangement had been made to the subject's satisfaction the experimenter said, "Kindly indicate the confidence you have in each rating you have made." Two or three, only, were troubled by this, asking if the exact rating was meant, to which the experimenter answered, "Why, yes, within a place or two." If a subject did not know one he was to rate by name, which occasionally happened, the card was placed to one side. Each one rated himself along with the others. In spite of the obvious difficulty of separating this trait from others and rating it by itself, very few showed any hesitancy in proceeding, though the middle part of the series caused more uncertainty, as a rule, than the ends.

#### *Method of Scoring:*

Although the different members of each group had been in frequent contact with each other for the better part of a year, it happened that some subjects could not call to mind some of the persons whose names appeared on the cards for them to rate. The usual explanation was, "I know all those fellows, but I don't know their names." Or, in the case of Groups II. and III. especially, such and such a person "doesn't come around when I'm here, I guess."

The result of this was that in Group I. 26 of the 225 ratings were missing; in Group II., 5 of the 225, and in Group III., 18 of the total 144. The danger here lies in the possibility that the tenth, say, in a list of only ten subjects would have had thirteenth place if all fifteen subjects had been listed.

In order to overcome this difficulty, the scores for Group I., where was the greatest number of omissions, were converted according to the Ream table for comparing incomplete order-of-merit ratings.<sup>1</sup>

<sup>1</sup>Ream, M. J.—A Statistical Method for Incomplete Order-of-merit Ratings. *J. of Appl. Psychol.*, 1921, 5; 261-266.

When the rating in the raw score was compared with that in the scores thus weighted, it was found that there was a correlation of .989 between the two. The only difference was that two names in the center were displaced two places and one name one place. In view of this close relationship, the more elaborate method was discarded as unnecessary in a case in which so few of the ratings were missing.

#### 4. SELF-ESTIMATES AND THOSE OF OTHERS

Apart from the relation of these results to the confidence scores of the test situations, which will be given later (IV., 1) the chief matter of interest in this connection is the relationship of the ratings made by others to those made by the subjects of themselves.

The subjects in Group I. tended to rate themselves higher than they were rated by others, the average rating of the group being 7.58, whereas the average of the ratings as each subject rated himself was 5.73, or nearly two places higher. This would seem to be in accord with other experimental findings.<sup>1</sup> But this does not carry through in the other groups, the two averages for Group II. being identical; and in Group III. we find the situation reversed.

TABLE I.

Group—	I.	II.	III.	Total.
Self-estimates higher .....	67%	53%	42%	55%
Self-estimates same .....	13%	14%	8%	12%
Self-estimates lower .....	20%	33%	50%	33%

Per cent of subjects rating themselves higher or lower than they were rated by others.

The above table shows this relationship. Perhaps young men consider it more of a virtue to be self-confident than do slightly more mature women; or perhaps the women just feel they haven't so much self-confidence, or act as if they are more self-confident than they feel!

The relationships between these two series of ratings are shown in the following table; the correlations are not high but are surprisingly uniform, being .44, .40, and .39 for Groups I., II. and III. respectively.

<sup>1</sup>Hollingsworth, H. L.—Judging Human Character, p. 48 ff.

TABLE II.

		Arith. M. of Ratings		Correlation	
		by Others.	by Self.	between the two.	
				R.	P. E.
Group	I.....	7.58	5.73	.44	.14
Group	II.....	7.89	7.60	.40	.15
Group	III.....	5.66	7.00	.39	.16

Relation of self-ratings to those of others in different groups.

## 5. DETAILED PROCEDURE

### TEST I.

#### Line Discrimination

##### *Materials:*

Ten cards<sup>1</sup> of white card board, 14 in. by 6 in. in size. In the center of each card a horizontal line was drawn, from the left end of which, facing the observer, a length of 100 mm. was cut off by an upright vertical line 5 mm. long and 2 mm. wide. This 100 mm. length served as a standard line for each card, while the remainder of the line (to the right of the vertical upright) varied in length from 95 mm. to 105 mm., and served as a comparison line to be judged longer or shorter in terms of the standard line. These lines were uniformly 1 mm. thick and were drawn in India ink by an expert draughtsman. The comparison lines were 95, 97, 98, 98.5, 99, 101, 101.5, 102, 103, 105 mm. long respectively. The cards were exposed in a dark room with uniform illumination, the subject sitting ten feet away.

##### *Directions:*

On each of these cards is a line divided into two parts in such a way that one part is longer or shorter than the other part, never equal. What you are to do is to compare the length of the two sections. The left hand part of the line is always constant. Your judgment, then, is as to whether the *right-hand section of the line is longer or shorter than the left*. You may inspect each card as long as you wish to make your judgment. As soon as each judgment is made, ascribe a degree of confidence to it (A, B, C, or D) according to the preceding directions.

<sup>1</sup>Garrett, H. E.—A Study of the Relation of Accuracy to Speed, Arch. of Psychol., 1922, 56, 52-53.

The series was run through twice, the second time in a different order, each order being a chance one, but kept constant throughout the experimentation.

Score: Number right minus the number wrong.

#### TEST II.

#### Weight Discrimination

##### *Materials:*

Eight standard weights, uniform in size and painted black, varying in weight as follows: 84, 88, 92, 96, 100, 104, 108, 112 grams. A black screen three feet square, was used to keep the weights from being seen by the subject.

##### *Directions:*

This is an experiment in lifting weights. The lifting will be without vision to cut down the effect of secondary criteria. In order to standardize the procedure somewhat, be sure to use the following method,—

1. Lift the weights with the same hand each time.
2. Lift them between the thumb and fingers.
3. You may heft each weight as long as you wish and as many times as you wish to make your judgment.
4. Your judgment is as to *whether the second weight is heavier or lighter than the first.*
5. When your judgment is made, endeavor to ascribe a degree of confidence to it, as you did in the preceding experiment.

The weights will be given you to lift in pairs,—a standard weight [which was the 100g. weight] and then one of the variables. The constant weight will thus be presented alternately with the others throughout the experiment.

The series of eight weights was run through three times with each subject, except that the 112-gram weight was not used the third time, thus making twenty judgments. The weights were presented in a chance order that was kept constant.

Score: Number right minus the number wrong.



TEST III.

Handwriting Comparison

*Materials:*

Twenty pairs of cards in twenty different hands, the inscription below appearing twice in each hand,—

Department of Psychology,  
Columbia University,  
New York City.

The cards were of uniform size,  $3\frac{1}{2} \times 5$  inches; none of the writing was done by any of the subjects to be tested. The inscription gives fairly complete data for comparison, since only six letters of the alphabet do not appear. One series of twenty handwriting samples was pasted in five rows in random order on a square of black paper muslin, so that they could be exposed more readily.

*Directions:*

For every sample of handwriting spread out before you, there is one to match it in the pack, though the numbers of the two sets bear no relation to each other. [These numbers were to facilitate recording.] Place each sample that is in your hand on its mate on the table beginning with Card One, and going right through the pack in order. As you match and place each card, ascribe a degree of confidence in your judgment as to whether it is correctly placed, using the following formula for convenience: "Card 27 belongs on card 22,—A (B, C, or D) Confidence." At any time a card that has been placed may be taken up and put down again, or another put in its place.

The score was the number rightly placed.

TEST IV.

Memory Span for Digits

*Directions:*

This is a test for memory span for digits. You will begin to repeat each number after it has all been given to you. In each case as soon as you have finished your repetition, say how confident you are that it was correct. *Your judgment, then, will be Right (or Wrong),—A (B, C, or D),* according to how confident you are that your repetition was correct or incorrect.

The digits were given orally at the rate of one a second.

There were twenty numbers, the first four with six digits each, the next four with seven digits, the next with eight, the next with nine, and the last four with ten digits each.

Score: The number of complete numbers repeated correctly. When a subject gave four successive A-wrong responses, the test was discontinued.

#### TEST V.

#### Performance

##### *Materials:*

Woodworth-Wells Number Blanks, Form A,<sup>1</sup> Columbia A Test Blanks,<sup>2</sup> a hand dynamometer and stop-watch.

Woodworth and Wells have found that the halves of their blanks are of equal difficulty, and they suggest that one-half of the blank is a sufficient test. For the purpose of this experiment, the blank was still further divided into four parts. The practical necessity for cutting down the time forced this procedure. Even if the four tasks are not of equal difficulty, though it would seem that they are, it would not materially affect the results of the experiment. Twenty-five digits, then, were to be crossed out in each case, five to a line, from amongst 250.

It was observed that no practice effect was evident, in part because of the distraction afforded by the subject's trying to better his record, which frequently resulted in his going back for a digit that he had skipped in his haste, and in part because the subjects had all done cancellation tests before.

The Columbia A Blank was treated in a similar fashion, being divided horizontally between the sixth and seventh lines, forming practically equal tasks with the last or thirteenth line eliminated.

Woodworth and Wells state that since just five digits were to be checked in each line, the errors on the number-checking test were so infrequent that they could be disregarded. This same result was achieved on the Columbia A test by indicating the number of A's to be crossed out at the end of each line. There were 233 and 229 characters respectively on each part of the blank from amongst which 46 were to be crossed out in each part. Thus the test, while keeping the procedure the same, made a slightly different experimental situation.

<sup>1</sup>Woodworth, R. S., and Wells, F. L.—Association Tests, Psychol. Monog. 1911, (No. 57) p. 24.

<sup>2</sup>Cattell, J. M., and Farrand, L.—Physical and Mental Measurements of Students of Columbia University, Psychol. Rev., 1896, 3, p. 641. Whitley, Mary T.—An Empirical Study of Certain Tests for Individual Differences, Arch. of Psychol. 1911, (No. 19) p. 61. Wissler, C.—The Correlation of Mental and Physical Tests, Psychol. Monog. 1901, 3, (No. 16).

The hand dynamometer was used for four trials, with a rest period of fifteen seconds between performances.

*Directions:*

The next test will be one in cancellation. It is the only one in which speed counts. Go across the page as rapidly as possible from left to right as in reading, crossing out all the 2's in every line. There are five 2's in each line, so be sure to cross out all five in each line before passing on to the next. Stop at the end of the fifth line, *i. e.*, when you get down to the first horizontal pencil line; the page will be broken up into four tasks with a slight intermission between each. Start when the signal is given, and be sure to let the experimenter know the moment you have finished.

The directions in the Columbia A test as outlined above, were given to the subject orally, as were those for the use of the hand dynamometer.

There were thus ten tasks, four Woodworth-Wells, two Columbia A, and four hand dynamometer. After the first cancellation trial and the first dynamometer trial, each subject was asked, "How do you think that performance compares with the average for college men (or women)?" After the other eight trials each subject was asked, "Do you think this performance was better or worse than the last?" After the last trial on each of the three tests, each subject was asked, "Suppose you had ten more trials, (in the case of the strength test the phrase was added, 'with sufficient time between to eliminate the fatigue factor') do you think some one of them would be better than one of these so far, or not?" With the other seven tests, each subject was asked, "Do you think you will be able to do better on the next trial?" That made two questions for each trial, or twenty questions in all, which called for the subject's evaluation of his confidence in his own work, past and to come.

The score was the number right minus the number wrong. Obviously, there was no standard for scoring the answers to the question, "supposing you had ten more trials . . ." These questions were made a part of the experiment as pertinent to the case, and of importance. In scoring for achievement, however, they were disregarded, being counted right in any case. If it had been practicable to do so, a complete test of this sort might have been prepared and scored without reference to possible outcome like tests XIII.-XVI.

To score the answers to the question, "Do you think this performance was better or worse than the last?" the performance of each test

was compared with that of the preceding. To ascertain which answers were right to the question, "How do you think this performance compares with the average?" it was necessary to procure norms for the different performances. Considerable difficulty was experienced in getting representative norms, particularly for the number-checking test. The following table shows the range available from which to choose, the numbers given being the average number of seconds it took to complete one-half the sheet. The present findings are included for comparison:

Woodworth and Wells <sup>1</sup> (Digit 0):	
20 men .....	72.5 sec.
20 women .....	61.5 sec.
Bingham <sup>2</sup> (Digit 1):	
200 men .....	48.3 sec.
Kitson <sup>3</sup> (Digit 6):	
31 men, 9 women (together) ..	86.87 sec.
Carothers <sup>4</sup> (Digit 3):	
200 women .....	77.64 sec.
Trow (Digit 2):	
30 men .....	67.86 sec.
12 women .....	64.62 sec.

---

<sup>1</sup>Op. cit.

<sup>2</sup>Bingham, M. V.—Some Norms of Dartmouth Freshmen, *J. of Ed. Psychol.*, 1916, 7, 131, 134.

<sup>3</sup>Kitson, H. D.—The Scientific Study of the College Student, *Psychol. Monog.* 1917, 23, (No. 98) 21-23.

The figure that Kitson gives, which is quoted by Carothers (See reference, foot note 4) page 35, as meaning seconds is, in reality, the average number of digits checked in two minutes! The test was thus altered that it might be given to a group. On page 22 Kitson says, "Each digit checked correctly counted one unit. No deductions were made for omissions or wrong figures checked." If Kitson's subjects averaged 69.2 digits in two minutes, we might infer that they checked 50 digits in 86.87 seconds. This is very slow time but might be explained first, because one of the most difficult digits was used, the digit 6. Second, because the test was given to a group, which was not the procedure of the other tests, apparently. Third, because it would seem that his subjects were not familiar with psychological tests. His preliminary directions were as follows: "I wish to quiet any fears you may entertain about these tests by assuring you that there is nothing mysterious or occult about them." Fourth, because of the unusual directions, which read, "Make any kind of a mark you wish. If you happen to make a mistake and cross out the wrong number, do not stop to erase,—simply draw a ring around that number and I will understand." There was a total of 55 errors made by 18 out of the 40 subjects. It may be supposed that the drawing of these 55 rings (if they were all drawn) in addition to the ingenuity in the kind of checks used, the novelty of the task, for it was the first of a series, together with the other considerations mentioned, might easily raise the norm ten seconds or so.

<sup>4</sup>Carothers, F. E.—Psychological Examinations of College Students, *Arch. of Psychol.*, 1921 (No. 46).

Possibly this wide range is due in part to the difference in the difficulty of the task depending upon which digit is checked.<sup>1</sup> This might help to explain the speed of Bingham's 200 men who checked the 1's in 48.3 seconds, which is considerably faster than Carother's 200 women, whose average was 76.64 seconds, in checking 3's.

Results: In the following table are set forth the results in terms of the Ar. M. of this experimentation as compared with the norms used.

TABLE III.

	Woodworth-Wells Number-checking, in seconds for ½ the sheet.		Columbia A, in seconds, rate for whole sheet		Strength of grip in kilograms	
	Av.	Norm. <sup>2</sup>	Av.	Norm. <sup>3</sup>	Av.	Norm. <sup>4</sup>
Men (30).....	67.86	72.5	88.89	100.	43.33	42.
Women (12).....	64.62	61.5	88.04	87.3	22.83	27.2

In strength of grip, the men come out a little above the average, while the women are considerably below it, although it is hardly fair to make invidious comparisons on the basis of twelve cases. In number-checking in both tests, the women run very close to the norm, whereas the men are much ahead of theirs, bringing the two groups very close together.

TEST VI.

Spelling

Materials:

A list of twenty common though rather difficult words, eleven spelled correctly, and the remainder with some slight error in the spelling.<sup>5</sup>

<sup>1</sup>Woodworth and Wells give the following:  
 Easiest .....1, 7.  
 Next .....0, 4.  
 Next .....2, 3, 5, 8.  
 Hardest .....6, 9.

<sup>2</sup>Woodworth-Wells, op. cit.

<sup>3</sup>Whitley, M. T.—Op. cit.

<sup>4</sup>These norms were obtained from the Columbia and Barnard Departments of Physical Education as the "norms for all colleges." For the women, this norm was checked with the measurement of strength of grip of 827 freshmen and seniors of the years 1921, 1922, 1923, for which figures were available, the average of which was 60.8 lbs., very close to the 60 lbs. (27.2 kg.) used.

<sup>5</sup>See appendix.

*Directions:*

In the space at the left of each word below, place an R if the word is rightly spelled, and a W if it is wrongly spelled. Just to the left of each W or R, place an A, B, C, or D to indicate the degree of confidence you have in each case. There may or may not be an even number of words that are right or wrong, so this criterion should be avoided. Let the experimenter know when you have finished to your satisfaction.

Score: The number right minus the number wrong.

## TEST VII.

## Incidental Memory

*Materials:*

A Blank with numbers from 1 to 20 down the left margin.

*Directions:*

Write on this sheet as many of the words used in the preceding test as you can remember. You need pay no attention to matters of arrangement or spelling, as those things are not scored. Record at the left of the numeral in each case, the confidence you have that the word written there was on the preceding list.

Three minutes were allowed for the task, though most subjects had reached their limit long before this.

Score: Number right.

## TEST VIII.

## Recognitive Memory

*Materials:*

A list of forty words, twenty of which, appearing at chance intervals, were the ones on the former list, except that the spelling was corrected. The remaining twenty<sup>1</sup> were with one or two exceptions the ones that came to the experimenter by free association from the words of the other list. They therefore bore some similarity to the original list either in meaning or sound or appearance, etc., and made the test more difficult than totally different words would have made it.

*Directions:*

In the space at the left of each word below, make a check mark if you think it appeared in the former list of words. Just to the left

---

<sup>1</sup>See appendix.

of each check mark, place an A, B, C or D to indicate the degree of confidence you have in each case.

Score: The number right minus the number wrong.

TEST IX.

Geographical Estimates of Size

*Materials:*

A blank calling for the following:

- (a) The five largest cities in the United States.
- (b) The five states in this country having the largest area.
- (c) The five states in this country having the smallest population.
- (d) The five states in this country with the largest population.

*Directions:*

List, in the spaces below, the data called for according to your best judgment, without regard to the order in which you enumerate them. To the left of each numeral record your confidence that the item listed properly belongs in that list.

The answers were found in the World Almanac; the score was the number right.

TEST X.

Logical Fallacies

*Materials:*

Blanks with a list of propositions,<sup>1</sup> some of which are fallacious and some not.

*Directions:*

In the space at the left of each statement place a minus sign if it contains a logical fallacy, and a plus sign if it contains no such fallacy. To the left of each plus or minus sign place an A, B, C, or D to indicate your confidence that the statement is fallacious, or consistent. Notice that this does not call for your belief or disbelief, but only your judgments in regard to the statements as given.

In practically every case, the fallacies were of the formal type. A number of them were obtained from Bradley's Logic.

Score: The number right minus the number wrong.

---

<sup>1</sup>See appendix.

## TEST XI.

## Addition

*Materials:*

A blank<sup>1</sup> on which were five columns of single-digit figures, four columns of two-digit figures, four columns of three-digit figures and three columns of four-digit figures, all with three figures to a column. The blank was so arranged that the sums of the single-digit figures were to be tabulated and their sum added to the sums of the two-digit figures, and so on, in order that any error anywhere in the list would appear in the final total.

*Directions:* (Given orally to avoid confusion)

Add the five columns of single-digit figures in Row 1, placing the sums underneath. . . . Transcribe the sums to positions under the first column as indicated by the arrows. . . . Add this column of sums and then the two-digit figures in Row 2. . . . Transcribe these and proceed as before until the final total is found. . . . Now, beginning with this last total and working up, ascribe a degree of confidence to each sum obtained. . . . Now check over your work, making any changes that you care to, and record your confidence in the sums as you then find it.

An arbitrary scoring system was used. One point was taken off for an error in transcription, two for an addition error the first time over, and four for such an error if it was not corrected in the check. An error was called such but once, though the mistake carried on down to other sums.

Score: Twenty (the number of sums) minus the number of points deducted.

## TEST XII.

## Ethical Judgments

*Materials:*

Blanks containing a list of mooted ethical questions.<sup>1</sup>

*Directions:*

Place a Y on the line before each question you would answer by Yes, and an N before each question you would answer by No. Place an A, B, C, or D just to the left of your answer in each case, to indicate your degree of confidence in it.

<sup>1</sup>See appendix.



Inasmuch as there is no objective reference by means of which the correctness of responses on this and the following tests could be ascertained, they were not scored for achievement, like those preceding, but only for confidence like all the tests, in the manner explained in Section 6.

## TEST XIII.

## Causal Judgments

*Materials:*

Blanks containing ten propositions or questions with four reasons given in support of each.<sup>1</sup> The reasons are all more or less applicable, and thus not like the "Test of Common Sense" of the Alpha Intelligence examination.

*Directions:*

Below is a series of questions with four answers given to each question. Indicate with X's in the left margin the *reason or reasons* in each case which you consider the most nearly right. Just to the left of each X, place an A, B, C, or D, to indicate your degree of confidence. If what you believe to be the real reason in any case is not given you may insert it.

## TEST XIV.

## Belief

*Materials:*

The Sumner list of twenty-five Beliefs.<sup>2</sup>

*Directions:*

In the left margin, on the line before each question, place a Y if your answer is Yes, and an N if your answer is No. Just to the left of your Y or N, place an A, B, C, or D, to indicate your degree of confidence in your answer.

## TEST XV.

## Judging Poetry

*Materials:*

Abbott-Trabue Poetry Judging Leaflet, Series X.<sup>3</sup>

<sup>1</sup>See appendix.

<sup>2</sup>Sumner, F. B.—A Statistical Study of Belief. Psychol. Rev., 1898, 5, 616-631.

<sup>3</sup>Abbott, A. and Trabue, M. R.—Exercises in Judging Poetry. Bureau of Publications, Teachers Col., Columbia Univ., 1921.

*Directions:*

In addition to the instructions which are given, evaluate your judgments of "Best" and "Worst" with the same A, B, C, D scale that you have been using to indicate your confidence in the correctness of your judgment. Let the experimenter know when you have finished judging the first ten sets of poems, one through ten. The others may be omitted. If, however, you are so familiar with some of the poems that your judgment of "Best" would be no more than recognizing the well-known originals, omit any such, and substitute in their stead from the three final sets beginning with eleven.

Although this arrangement brought it about that in some cases one or two different poems were judged, it was believed that this possibility of error was not so great as that of ascribing an A-Confidence for æsthetic judgment, when the mental operation performed was recognition.

The ten Worst could not well be scored, since there is no standard for worst, but the ten Bests could be checked up against the originals.

Score: The number right multiplied by two, to get it on a basis of twenty judgments.

## 6. THE METHOD OF SCORING CONFIDENCE

In order to get the confidence score, all the A-judgments made by any one subject were totaled and multiplied by four, the B-judgments by three, the C-judgments by two, and the D-judgments by one. It is recognized that this was a purely arbitrary method of proceeding, but seems admissable for a number of reasons.

It is desirable to have some one score that shall represent the subject's confidence in the different situations of the experiment, in order that he may be ranked and that relationships may be sought with other measures. Four measures are too bulky to handle in this way.

Granted, then, that some one measure is desirable which shall comprehend the four, such a measure must give an A-judgment more weight than a B, a B than a C, and a C than a D. For example, if one subject should give twenty A-judgments, and another twenty D-judgments, obviously the one giving the A-judgments is more confident in those situations. But how much more confident he is, it is impossible to say. It will readily be admitted that the use of four degrees of confidence furnishes no sufficient ground for the conclu-

sion that he is four times as confident. However, for the purposes of scoring, this seems as satisfactory a scheme as any.

It was thought that possibly a weighting of each degree of confidence in accordance with the frequency with which that confidence appears, might be more advisable, despite the fact that this would assume a probability curve.<sup>1</sup> But there is no reason for thinking that because there is a greater number of A-confidences that each one is worth less. Each is an absolute measure of an individual's subjective state, no matter how often that state may recur. The verbal response often is, "Well, I'm absolutely positive of that," in the ascription of an A-confidence. The same thing is true of the other stages.

As positive evidence in support of the scoring device, let us consider the relationship of the total confidence scores to the A- and to the D-judgments.

We should naturally expect that persons who are often absolutely confident and who are rarely uncertain would stand high in confidence. If a scoring device were employed using these criteria only, it would obviously be unsatisfactory because it omits all consideration of the B- and C-judgments. But on the other hand, any plan used should correlate well with these criteria, for if it did not the plan used would have something fundamentally wrong with it. As a matter of fact, when the forty-two subjects were ranked according to their total confidence scores, on a basis of the number of A-judgments each made, from the greatest to the smallest, and to the number of D-judgments each made from the smallest to the greatest, it was found that the correlation of the confidence scores with the frequency of A-judgments was .83, and with the infrequency of D-judgments was .82; P. E. .03.

On the basis of the above considerations, then, the weighted score above discussed was used as a measure of confidence.

Most of the test situations gave opportunity for twenty judgments. This was not the case, owing to their inherent nature, with tests IV., VII., XIII., XIV., XV. and XVI. To make the confidence scores comparable, therefore, they were weighted for these tests on the basis of twenty, according to the following formula:  $\frac{20}{N} \times C$ , where N equals the total number of judgments made for any one test, 18, 25, etc., and C is the number of judgments for each degree of confidence.

<sup>1</sup>Thorndike, E. L.—*Mental and Social Measurements*, Table 22, p. 117. Table for the transmutation of measures by relative position into measures in units of amount.

## III.

## CONFIDENCE AND ACHIEVEMENT

## THEIR DISTRIBUTION AND CORRELATION

The nature of the two distributions, that of the achievement scores and the confidence scores is shown in Table IV. This table was derived from a tabulation of the scores made by each subject on Tests I. to XII. This limit was set because the remaining tests were without any objective check, and so could not be scored for achievement.

The upper two-thirds of the table, then, indicates the nature of the two distributions, which show a certain amount of similarity. The quartile deviations with their wide dispersion of measures are more alike than appears at once, as is indicated when the absolute measure is translated into one of relative variability by the following formula:

$$\text{Coefficient of Variation, } V = \frac{Q}{\text{Median}}$$

Likewise there is little difference to be noted between the three groups.

With such a distribution, therefore, the resulting correlation (Table V.) is all the more striking. Using the rank-difference method, it is found that for all subjects the correlation between achievement scores on the first twelve tests, the scorable ones, and confidence scores on these same tests is  $-.03$ ; P. E., 10.

TABLE IV.

		Range	Mean	Median	Q.	V.
<i>Achievement</i> scores on scorable tests. I-XII.	Group I ...	168-114	140.9	138.0	17.5	.13
	Group II ...	184-103	138.5	135.0	15.0	.11
	Group III ...	155- 94	129.8	135.5	21.5	.16
	Total .....	184- 94	136.9	136.5	14.5	.12
<i>Confidence</i> scores on scorable tests. I-XII.	Group I ...	817-692	772.5	791.0	24.5	.03
	Group II ...	873-621	768.6	792.0	55.5	.07
	Group III ...	844-629	737.0	745.0	25.5	.03
	Total .....	873-621	761.0	779.0	41.5	.05
<i>Confidence</i> scores on non-scor- able tests. XIII-XVI.	Group I ...	273-184	238.8	249.0	15.0	.06
	Group II ...	267-185	214.3	211.0	18.0	.09
	Group III ...	276-136	229.7	235.0	28.5	.12
	Total .....	276-136	227.4	230.0	25.0	.11

The Confidence and the Achievement distributions for the three Groups and the total Group.

This coefficient indicates that there is practically no relationship between the confidence scores of the subjects on the twelve tests and their correctness in the performance of those tests. From this it is perhaps not impossible to infer that there is little or no relationship between people's rightness and their general confidence; that they are not necessarily generally right if they are generally confident, and *vice versa*.

This does not mean, as another part of this study clearly shows, that a person is not more apt to be right if he is highly confident, for he is. What it does mean is that the people who tend to be more confident than others are as likely to be right as the unconfident; and the unconfident are as likely to be right as the confident. If a person is generally assertive, he is no more, or less, apt to be right than a person who is generally not assertive. This seems quite reasonable, and consonant with common experience; but it is an easy thing to lose sight of, for instance, while listening to a salesman, perhaps, or a politician.

The lower third of Table IV. shows a similar homogeneity, but yields a very different correlation when comparison is made with the confidence scores of the scorable tests. For it is pertinent to ask if the subjects who were confident when there was a possible correct answer, an objective standard, are also the ones who are confident in mere matters of opinion or belief.

Here we have a very curious result. The correlation in this case is .54. This distribution shows that some were confident to about the same extent in both types of situations; others were confident when it came to evaluating the results of their own intellectual labors, but more dubious in matters of opinion, while still others were cock-sure in matters of opinion or belief, but quite uncertain of their own results in addition, discovering logical fallacies, and the like. It does not mean that they are divided off into types, for all sorts of intermediate cases are found.

In spite of the positive nature of the correlation, one subject ranks fifth in confidence in Tests I.-XII. and thirty-sixth when there is no objective check; while another ranks thirtieth in the former and fourth in the latter.

It is an interesting coincidence, if it is nothing more, that the two subjects in Group II. with the greatest range, 31 and 25 respectively, both showing very high confidence in the scorable tasks and very low in the others, are very often together, the one who is the older having selected the younger to be his laboratory assistant. In Group III.

the two subjects having the greatest range, this time with low confidence in the scorable tests and high in the others, are fast friends and boon companions.

The correlations above discussed may be tabulated by groups as follows :

TABLE V.

	GROUP I		GROUP II		GROUP III		TOTAL	
	r.	P. E.	r.	P. E.	r.	P. E.	r.	P. E.
Correlation between achievement and confidence scores	-.15	.17	.22	.17	-.15	.17	-.03	.10
Correlation between the two kinds of confidence scores based on tests I-XII and on XIII-XVII . . . . .	.86	.04	.75	.07	.25	.18	.54	.07

It is perhaps significant that the younger group of men was more consistent in the matter of confidence in the two kinds of situations (.86), and that the women were considerably below the others in this respect (.25).

We seem to have two different kinds of situations here, such that the same group of people line up differently in them. Yet it is all confidence. It seems from these results that it is not quite safe to talk about the confident person, or the confident type of person; for it would surely mean that he would tend to be confident in both these types of situations, and we have seen that there is an even chance that he would not, and if the subject is a woman, we might guess that the chance would be less than even.

We shall have occasion to recur to this point.

## IV.

## OTHER RELATIONSHIPS

## 1. CONFIDENCE SCORES AND RATINGS

It is next of interest to inquire how nearly the arrangement of names, ranked according to the total confidence scores, compares with the arrangement ranked according to the ratings for confidence. In other words, what is the relation of confidence as measured by these test situations to confidence as measured by the opinions of acquaintances?

This relationship is shown in the second of the three columns in Table VII. Here it is observed that the highest correlation obtained between the confidence scores and the order-of-merit ratings for confidence was .37 and the lowest .16, with P. E.'s so high as to make the correlation of little real value. Indeed, the correlations are so low that they might easily lead one to suppose that they were between entirely different traits; and I am inclined to believe that this is the case.

Turning for a moment from the general results, let us compare the individual subjects in this particular. When the total confidence score of each subject was compared with the average of the ratings he was given, the two being made comparable by ranking, it was discovered (Table VI.) that in rare cases a subject had the same place in the list, first, fourth, seventh, etc., in both rank orders, like Subject Su in Group I.; whereas in others there was a divergence of as much as ten places in the two lists for the same subject, with but fifteen individuals ranked: *c. g.*, Subject Sm. in Group I. was ranked fifteenth while his confidence score placed him fourth.

Furthermore, this divergence is in both directions: In Group II., for example, subjects An, Jh and Mi are rated as having much less confidence than the tests gave them, while subjects Ad, Ho, Kl and Le are rated with more. The former, who were rated low in confidence but scored high, are pleasant-mannered, agreeable men whose opinions are presented when asked, but not otherwise, usually, and then without in any sense forcing them. The others, who were rated high in confidence but scored low, are fond of argument, even though they may realize that their position is untenable; or they have a certain reserve of bearing; or, an exact knowledge about certain things which, when asked for, is given with a high degree of assur-

ance; and it is therefore assumed by their acquaintances that this same degree is maintained in other situations as well. It seems to be clear that these other personal characteristics go into the rating.

In other words, the ratings are subject to at least two decided fallacies. First, the subjects are rated for some other trait or traits than confidence, though the judges are perfectly conscientious about their ratings. Second, the ratings are based on too few situations in which the appearance of the trait might be manifest to the rater.

I believe it is safe to say that the same conditions hold for the self-ratings, though, perhaps, to different degrees.

TABLE VI.

GROUP I				GROUP II				GROUP III			
Subj.	Oth. Ra.	C. Sc.	Self Ra.	Subj.	Oth. Ra.	C. Sc.	Self Ra.	Subj.	Oth. Ra.	C. Sc.	Self Ra.
Ca	7	1	4	Ad	2	7	2	Ar	9	8	5
Eh	4	5	4	An	10	1	9	Be	5	7	5
Gi	11	9	9	Br	1	2	9	Bu	2	2	5
Mo	2	3	2	Cr	7	10	5	Ch	8	11	10
Oc	8	11	11	Gr	12	8	6	Do	10	6	6
Pa	3	6	7	Ho	3	11	4	Gi	11	4	9
Ph	5	13	2	Jh	11	3	6	He	4	12	6
Pr	1	7	5	Jo	4	4	12	Hv	6	5	10
Ri	13	15	9	Kl	5	13	6	Hu	1	1	4
Sc	95	8	2	Lk	8	14	4	Le	12	9	8
Sh	12	12	7	Ma	9	6	2	St	7	3	6
Sm	15	4	4	Mi	14.5	5	11	Wi	3	10	10
Sp	95	2	8	Ne	6	9	12				
Su	14	14	8	Ni	14.5	12	13				
Ti	6	10	4	We	13	15	13				

The ranking of the subjects of the three groups, according to the Confidence Scores and Ratings.

TABLE VII.

	*Self-ratings and Those of Others		Others' Ratings and Confidence Scores		Self-ratings and Confidence Scores	
	r.	P.E.	r.	P.E.	r.	P.E.
Group I.....	.44	.14	.37	.15	.38	.15
Group II.....	.40	.15	.16	.17	.00	.17
Group III.....	.39	.16	.25	.16	.50	.15

\*Repeated from Table II.

Correlations of Confidence Scores with Ratings.



## 2. CONFIDENCE AND ACHIEVEMENT SCORES AND INTELLIGENCE

It was possible to secure intelligence ratings for Group I. only; but here the results are sufficiently clear to indicate that what we are dealing with is something quite other than general intelligence as measured by the Army Alpha and the Thorndike Tests. In the latter, one person's score was missing unavoidably, so the score of that subject was omitted in the correlation.

The correlations of confidence and achievement scores with intelligence are as follows:

	r.	P. E.
Confidence and Army Alpha.....	-.42	.14
Confidence and Thorndike Intelligence.....	-.56	.12
Achievement and Army Alpha.....	.63	.10
Achievement and Thorndike Intelligence...	.64	.10

The negative correlations, suggesting the conclusion that there is an inverse relationship between confidence and intelligence, are certainly surprising, and, if borne out by further studies, present interesting possibilities for speculation.

So far as actual achievement on the materials of the test is concerned, as would be expected there is a marked positive relationship, for the situations are for the most part of the intelligence-test type. The relations we should not expect to find larger for the reason that some situations usually found on intelligence tests are missing, while others are added. Furthermore, the tests were not time-limit affairs, but each subject was given as long as he wished to finish.

## 3. COMPARATIVE VALUE OF THE SEPARATE MEASURES

There is always considerable interest in knowing which tests are the most valuable; but it is always necessary to assume that the standard or criterion with which they are compared, is a reliable standard. Here, the ones that suggest themselves are (1) the total confidence score, which, of course, is made up of the scores of the different tests, and (2) the ratings which were discussed rather disparagingly in section 1, above. The correlations, however, derived by the rank method, together with the P. E.'s in each case appear in Table VIII.

TABLE VIII.

Test	Correlation with Confidence Score		Correlations with Ratings					
	r.	P. E.	Group I		Group II		Group III	
	r.	P. E.	r.	P. E.	r.	P. E.	r.	P. E.
I—Lines .....	.76	.04	.69	.09	.70	.09	.68	.10
II—Weights .....	.72	.05	.84	.05	.64	.10	.81	.07
III—Writing .....	.23	.10	.19	.17	.43	.14	.44	.15
IV—Digits .....	.52	.07	.02	.18	.76	.07	.48	.15
V—Performance .....	.70	.05	.75	.08	.81	.06	.64	.11
VI—Spelling .....	.54	.07	.21	.17	.54	.12	.66	.11
VII—Memory .....	.29	.09	.41	.15	.48	.13	.08	.21
VIII—Recognition ....	.29	.09	.17	.17	.15	.17	.31	.18
IX—Geography .....	.50	.07	.52	.13	.51	.13	.48	.15
X—Logic .....	.38	.08	.60	.11	.30	.15	.31	.17
XI—Addition .....	.54	.07	.36	.15	.62	.11	.60	.12
XII—Ethical Judgments	.68	.06	.89	.04	.66	.10	.54	.13
XIII—Causal Judgm'ts	.64	.06	.60	.11	.74	.08	.35	.17
XIV—Belief .....	.56	.07	.78	.07	.60	.11	.33	.17
XV—Poetry .....	.41	.09	.52	.13	.36	.15	.48	.15
XVI—Rating .....	.50	.08	.62	.11	.78	.07	.44	.15

Correlations of the confidence score on the different tests with the total confidence score and with the ratings.

Tests I., II., V., XII. and XIII., judged by the standard of the total confidence score, seem the most reliable; III., VII. and VIII. the least. Since the ratings were by groups, it was necessary to figure the correlations in the same way; hence, in part because of the smaller numbers taken, the correlations are more erratic. Here Tests I., V., II. and IX. run high pretty consistently, while VIII., III. and X are low. If it seems desirable for any reason to give only a part of the series, selection might be made on this basis. In addition, it should be said that Tests III., VII. and VIII. are less valuable because of the lack of scatter in the results, both achievement and confidence being regularly high and fairly uniform for nearly all subjects.

## V.

CORRECTNESS AND THE DEGREES  
OF CONFIDENCE

"Oh, yes, I'm sure of that," says a friend of yours. How much can you count on his being right if he is so sure? Or, again, you are uncertain of something, so you give a guess; what are the chances of being right? It is this type of thing that the following portion of the investigation made an effort to find out.

This phase of the question of confidence has been touched before, but only with comparatively few subjects and limited kinds of material. The work of Fullerton and Cattell (see bibliography) is of course, classic. Hollingworth<sup>1</sup> found that subjects in evaluating their performance in tapping, color-naming and opposites tests, as "better than usual" or "worse than usual," were correct 98% of the time if their confidence was A, 81% of the time if it was B, 73% if it was C, and 59% if it was D. In another experiment the result was similar, 92, 73, 63, 60, for the A, B, C, and D respectively. Henmon's subjects<sup>2</sup> in judging the longer or shorter of two lines, tended to run lower, 91, 75, 59, and 41 per cent right respectively for the four degrees of confidence.

Two questions present themselves. First, How does correctness vary with confidence? Second, rising immediately from the first, How is this correctness affected by the type of situation?

In general, we can say in answer to the first question, using the data furnished by Table IX, that if a person is absolutely certain, he will be correct about 90% of the time. If he is fairly confident, he will be right about 75% of the time. If he is only slightly certain, he will be right about 60% of the time; and if his answer is a mere guess, he will tend to be right half of the time, with the chances slightly in his favor.

It might be urged that a distinction should be made here between the judgments which are of the Right-minus-Wrong type, that is, those of which half would tend to be answered correctly by chance, as in Tests I., II., V., VI., VIII., and X., and the others in which this is not the case. It might seem that in these latter, for example,

---

<sup>1</sup>Hollingworth, H. L.—Experimental Studies in Judgment. *Arch. of Psychol.*, 29, pp. 14 and 37.

<sup>2</sup>Henmon, V. A. C.—The Relation of the Time of a Judgment to Its Accuracy. *Psychol. Rev.*, 1911, 18, p. 199.

a mere guess would be correct far less than half the times. This will be shown to be true in certain cases, notably in Text IX., the questions on geographical size, in which only 25% of the D-judgments are correct. However, the average of all these latter is 91.3, 72.4, 52.1 and 51.0 per cent right for the A, B, C, and D groups respectively. This is not so very different from the whole group, as shown above, or from the Right-minus-Wrong alone, which runs 86.8, 71.9, 63.7, 53.0 per cent right for the A, B, C, and D groups respectively.

The great difference, however, lies in the deviations. The range was calculated for each by averaging the two highest and the two lowest figures representing the per cent correct for each degree of confidence; and as a result, it was found that whereas the average range so derived was 16.8 for all degrees of confidence in the Right-minus-Wrong set, it was 35.5 in the other.

We might fairly conclude from these data, therefore, that in the long run, the per cent of correct judgments for each degree of confidence remains the same whether the questions are of the Right-minus-Wrong type or not; but if they are not, a much greater irregularity appears.

These results are based on the first twelve situations, in which there is an objective check. The figures for the D-judgments are based on 860 judgments, as compared, say, with the A, which are based on 5,037; so they are not so reliable, presumably. The reason for the fewer cases seems to be that the type of material presented, together with the natural inclination of a subject to assume confidence in his own intellectual conclusions, threw a decreasing number of judgments into the B, C, and D columns respectively.<sup>1</sup>

It will be noted from the Q's that the A- and after that the D-judgments tend to present less scatter. This is natural when it is considered that these points are introspectively more easily accessible than the others.

So much for the final totals. We have answered the first question, namely, How does correctness vary with confidence? But how incomplete our answer is will be seen when we glance at the data more closely, with the other question in mind: How is this correctness affected by the type of the situation?

---

<sup>1</sup>The per cent of right judgments was derived from the totals, thus presenting a more exact figure than the arithmetic means of the averages, since the decimals were carried only to one place, and since the per cents in the D-column, being based on a smaller number of cases, tended to be somewhat non-representative. However, the medians of the percentage for each test situation deviate but slightly from the figures representing the per cent of right judgments.

It will be noted, as one lets one's eyes follow down the A-column in Table IX., that the Incidental and Recognitive Memory tests, for example, run above the average, while the Geography, Logic, and Poetry judgments run considerably below. This would mean that immediate recall and recognition are very dependable. Though not half the material is recalled, and while nearly all is recognized, that which is confidently recalled or recognized soon after presentation is practically certain to be correct. Notice, however, that in the C- and D-columns, these types of judgment tend to be far below the average, even less dependable than most.

If the geography questions are fairly typical of many of the half-estimate, half-memory type of questions that often confront us in every-day life, and I believe that they are, we may say of such that a person, when confident of his correctness, is not so apt to be right (80.9%) as in some other situations. This is true, too, in the evaluation of poetry, and in conclusions based on the reasoning processes.

TABLE IX.

	A	B	C	D
I—Lines .....	93.1	76.6	71.8	71.6
II—Weights .....	91.8	79.6	65.0	58.3
III—Writing .....	98.4	93.6	79.3	58.3
IV—Digits .....	96.7	79.5	63.3	51.5
V—Performance .....	83.1	73.1	68.8	55.5
VI—Spelling .....	82.6	64.8	57.7	49.1
VII—Memory .....	96.0	80.8	18.2	...
VIII—Recognition .....	96.4	75.2	57.3	40.6
IX—Geography .....	80.9	42.2	34.7	25.8
X—Logic .....	73.6	62.0	61.9	43.1
XI—Addition .....	93.0	84.2	66.7	82.7
XII—Poetry .....	83.1	54.3	50.5	36.7
Total Number of Judgments.....	5037	2470	1713	860
Number of Right Judgments.....	4569	1802	1055	464
Percent of Right Judgments.....	90.70	72.95	61.58	53.95
Median percent .....	92.4	75.9	62.6	51.5
Quartile Deviation .....	13.8	18.8	18.3	16.5

Per cent of correct judgments in each test for each degree of confidence for all forty-two subjects.

It will be noted that this unreliability continues right through the four degrees of confidence; so that in these situations, a mere guess has very small chance of being right.

Here, the question might be raised, Do the various functions have an equal chance? Perhaps the Geography Test is a great deal more difficult than the others, and that is the reason that only 80.9% of the A-judgments are correct. It must be remembered, however, that the 80.9% does not represent the per cent of questions answered cor-

rectly, or anything of that sort. It is not a score. Instead, it represents the per cent of correctness in the answers about which the subjects were positive they were right. Now, suppose the test be made a great deal more difficult. Obviously, the number of A-judgments would decrease, but it would seem reasonable to suppose that the per cent of them which were correct would remain substantially the same. In like manner, the number of D-judgments would increase; but it would seem reasonable to suppose that the per cent of them which were correct would remain substantially the same. Likewise, the number of B-judgments would decrease, and the number of C-judgments would increase, but to a lesser degree. Clearly the present investigation does not deal definitely with this problem, one which might afford interest for a later investigation. For the present, therefore, we shall assume that, unless the task is so simple as to throw all or nearly all the answers into the A-column, the figures would remain substantially the same with tasks of the same type but varying in difficulty, and that the functions are therefore comparable.

A few comments should be made about some of the resulting measures. The high per cent of D-judgments that are right in the Addition Test is based on too few individuals to be indicative. Nearly all subjects were more than D-confident in the correctness of their sums, though they had little reason for being! It is probable that the high degree of confidence in the Hand Writing Comparison Test was due to too great ease in its performance, though some found it very difficult and were not at all certain about their results. It is interesting to note that sensory discrimination as measured by the judgment of lines, and also, though to a less degree by the weights, runs toward an accuracy considerably higher than the average in the D-judgments. We have suggested that the reason for this is possibly not that they offer a 50-50 chance; other such judgments run lower. At least, other factors are in part responsible; the process is simpler than the others with less evasive criteria; also, uncertainty or even professed ignorance is not the reflection upon the subject that it is considered in the more strictly intellectual processes!

There were found to be no differences in the three groups of subjects that were at all significant. They were characterized, rather, by an interesting uniformity. As judged by the number of A-judgments in Groups I. and II., the younger men (Group II.) did not seem more confident, more youthfully cocksure than the older. It may be, however, that the older are not their seniors by a sufficient number of years to make such a difference manifest. The women

had fewer A-judgments in proportion to the number of cases than the men, and had a decidedly larger number of guesses. This may or may not be significant.

VI.

CONFIDENCE AS A CHARACTER TRAIT

We now come to the very heart of our problem. It is clear that the total confidence scores of the forty-two subjects can be placed in rank order, the highest at the top and the lowest at the bottom. It is an easy generalization to make, then, that those at the head of the list are the most confident, and the subjects become less and less confident as we go down toward the bottom of the list.

But can we say this? Is it fair to the data, to the totals, to use them in this fashion?

We can but admit that here is something positive. The upper four subjects scored 1,075 more points in confidence than the lower four, or nearly a quarter again as many. And we have seen that in the large, knowledge of the material, if this can be inferred from the degree of correctness of the answers, had no influence in the matter. (Correlation  $-.03$  between Confidence and Achievement Scores, Table V.).

We might even conclude that the cumbrous number of situations here employed is far too great. If confidence is a trait, find a test to test it, and let it be one that can be easily administered, by the group method preferably.

TABLE X.

Subj.	Range	Q.	Subj.	Range	Q.	Subj.	Range	Q.
Ca	1-35	6.7	Ad	2-42	6.8	Ar	5-42	11.1
Eh	1-37	9.6	An	1-33.5	6.9	Be	3.5-41	11.0
Gi	1-36.5	10.4	Br	5-27	4.6	Bu	1-37	9.9
Mo	3.5-27	3.5	Cr	9.5-38.5	6.1	Ch	14-41	5.6
Od	9.5-34	6.0	Gr	1-37	11.8	Do	1-42	8.8
Pa	2.5-35.5	8.9	Ho	11-41	5.0	Gi	1-42	13.1
Ph	11-40	7.5	Jh	2-40	10.6	Hd	5.5-42	6.5
Pr	4.5-35	7.6	Jo	3-34	9.9	He	3-39	9.4
Ri	11-41	6.9	Kl	14-41	6.6	Hu	1-41	9.8
Sc	2.5-34	5.8	Lk	2-41	6.9	Le	5.5-41	7.8
Sh	5-42	10.0	Ma	2.5-32	9.9	St	4.5-36.5	10.8
Sm	2-29	3.9	Mi	2.5-38	5.1	Wi	4-42	14.3
Sp	4-33	6.0	Ne	6-42	5.6			
Su	13-40	9.8	Ni	13-40.5	8.5			
Ti	1-34	7.5	We	14-42	3.8			

Range and deviation in the rank order of Confidence Scores on the different tests for each subject.

TABLE XI.

	Ar. M.	Med.	Q.
Group I.....	30.7	30.5	3.5
Group II.....	32.1	31.0	4.0
Group III.....	36.5	36.8	1.6
Total .....	32.8	33.0	3.8

Distribution of the ranges for the different groups of subjects.

TABLE XII.

	Ar. M.	Med.	Range	Q.
Group I.....	9.8	7.5	3.5-10.4	1.8
Group II.....	7.2	6.8	3.8-11.8	2.4
Group III.....	7.3	9.8	5.6-14.3	1.4
Total .....	8.0	7.6	3.5-14.3	3.9

Distribution of the Q's for the different groups of subjects.

Before we allow our enthusiasm to carry us too far, however, let us examine our data more carefully. The ranking<sup>1</sup> of the different subjects in the different test situations was derived by totaling the confidence scores for each subject in each test and placing the subject with the largest score first, the next second, etc. When more than one subject received the same confidence score on any one test, the median rank was assigned to each.

The range and deviation of these positions on the different tests are shown in Table X. For example, subject Ca in the sixteen tests was successively 14th, 8th, 20th, 35th, 11th, 17th, 13th, 19th, 5th, 4th, 11th, 1st, 7th, 2nd, 4th and 24th! His range, then, was from 1 to 35, or 34 places.

With such a range any single test, or even any total, or any average is fictitious, and has little or no reliability as an indication of any general quality or trait of confidence.

Even subject An, the most confident according to the total scores was 29th in Test VIII. and 33rd in Test XV., while two subjects, Do and Gi, placed first in one test and 42nd in another! Such a range, though the most extreme possible, is not so very much greater than the average (Ar. M.), which is 32.8 places, the median being 33. That is, on the average, the subjects were over 30 places apart in their highest and lowest confidence. The greatest range possible, 41, was mentioned above; the smallest found was that of Br, whose position on Test VI. was 5th and Test X. was 27th, or a difference of 22 places.

<sup>1</sup>The complete data for this as well as the rest of the experimentation will eventually be placed on file in the laboratory of Columbia University.



There is an indication of a sex difference in this matter of variation, for the average for the women is between four and five points greater than for the nearest male group; and their consistency is likewise greater, as indicated by the Q's in the summary of Table XI.

The quartile deviations of the rank positions of the different subjects on each test show considerable variation, running all the way from 3.5 to 14.3. For a distribution of only 42 subjects, this is, of course, excessive, the average being 8 (Ar. M. 8; Med. 7.6), Table XII.

For a further appreciation of the variability of an individual's confidence in different situations, we see from Table X that of the ten individuals who placed first in some one test, each one being more confident than all the other subjects in that situation, there was not one but ranked 33rd or lower in some other situation. Furthermore, of the nine subjects who were 42nd or least confident of all in some one test situation, there was only one who did not rank sixth or higher in some other.

Any test of confidence, then, which does not cover a wide number of situations is futile, as an indicator, and its predictions will as likely be wrong as right. It seems possible that this is also true of other character traits. The traits are not constants, but vary with the varying situations.

## VII.

## CONCLUSIONS

I. *Confidence and Achievement.*

1. The individuals who tended to be most confident, as judged by the total confidence scores, were no more apt to be right than the others. ( $r. = -.03$ .)

2. There was a correlation of  $+ .54$  between the confidence scores in tests scorable as right or wrong and the others.

II. *Confidence, Ratings, and Intelligence.*

3. Persons in rating themselves agreed moderately well with others. ( $r. = .41$ .)

4. Correlations between ratings and confidence scores were low.

5. For fifteen subjects, there was a negative correlation between confidence and intelligence scores.

6. A decided positive relationship was found between achievement and intelligence scores.

III. *Correctness and the Degrees of Confidence.*

7. In general, correctness varied directly with the degree of confidence. That is, the judgments which were given with a high degree of confidence, for example, were more apt to be right than the others.

8. The type of situation, however, greatly affected this per cent of correctness for the different degrees of confidence.

IV. *Confidence as a Character Trait.*

9. The range of each subject's confidence was so great in the different situations that any one test is necessarily fictitious as an indicator of any general quality or trait of confidence.

\* \* \* \* \*

The significance of this very definite conclusion cannot be over-emphasized. At the present time, when tests of all sorts are very much in vogue, it is no little temptation to seek a name and fortune by inventing a few and putting them on the market. The number of important things not measured by the current tests creates a strong demand, particularly in such of the applied fields as the selection of

employees; and it is here that the market is too apt to be glutted in a short time by a lot of useless if not harmful contraptions.

The business executive, the personnel man, the school superintendent want to select employees who are "trustworthy," "energetic," "industrious," who have "self-confidence," and "initiative," etc. They desire some means that is more dependable than the necessarily faulty estimate of others.

These are legitimate demands which are being voiced, and it is unfortunate that psychology has so little to offer in the way either of definite scientific knowledge, or of practical assistance through the medium of tests and measures. In view of the demand for the latter, it is to be hoped that they will not be forthcoming before the former. Probably the two will develop side by side in the virgin soil which the scientific study of character traits offers to the inquiring experimenter. In the process of working this soil, the psychologist will naturally cast a furtive eye in the direction of the intelligence tests, which have furnished such an abundant yield in recent years. He will want to avoid a certain excess of enthusiasm which has unfortunately characterized the earlier movement. He will study the situations carefully, so as not to be misled by any superficial parallelism.

In these intelligence tests, a number of situations are presented to each subject, who may be poor in one situation, average in another, and excellent in a third, and so on. His score is determined on the basis of the number of his correct responses.

It is natural to hope that a confidence score can be similarly determined. Perhaps a refinement of the measures here used will ultimately result in an approximation in this direction, reducing the divergencies recorded in Table X to something more nearly like those of a standard intelligence test. If this should be done, and a sufficient number of test situations used, something that might be called a measure of "general confidence" would result that might be useful for prognosis. It seems probable that some of the other phases of confidence as they are outlined in Section 2 of Chapter I might likewise be tested, and possibly the results thus gained might correlate more highly with ratings, if this is desired, and might be more constant in one individual, thus giving a test a higher prognostic value.

So far as confidence of judgment is concerned, however, apart from these other phases, the conclusion seems clear that the amount of confidence a person has in the correctness of his judgments is not constant, but varies widely from situation to situation; hence, any single test, judgment or reaction, or even a handful of them, is futile as a measure. It is conceivable that the same is true of the other so-called character traits.

## APPENDIX A

## THE INDICATORS USED

## SPELLING LIST, TEST VI

anoint  
 succesful  
 sacrilegious  
 delerious  
 speech  
 existence  
 geneology  
 dyspepsia  
 inoculate  
 announce  
 pervaricate  
 beneficent  
 caterpillar  
 operetic  
 surreptitious  
 interceed  
 persue  
 resistance  
 suppress  
 dispair

## RECOGNITION LIST. TEST VIII

lubricate  
 teaspoonful  
 religious  
 animate  
 speak  
 penitence  
 geology  
 dissertation  
 innocuous  
 propound  
 supplicate  
 malevolent  
 cocoon  
 theatrical  
 perennial  
 supersede  
 peruse  
 persistence  
 surprise  
 desperate

## TEST X—LOGICAL FALLACIES

- ..... 1. Men who succeed are always men of perseverance; therefore, if you would succeed, persevere.
- ..... 2. He must be a Mohammedan, for all Mohammedans hold these views.
- ..... 3. No A is B; all C is A; therefore no B is C.
- ..... 4. This battleship is one of the best of its kind, for it belongs to the best navy in the world.
- ..... 5. If he fails, it will be because he has not worked hard; and he has not worked hard; therefore he will fail.
- ..... 6. D equals C; B is greater than A; C is less than A; therefore B is greater than D.
- ..... 7. The League of Nations should be adopted, for shifting alliances, secret treaties, balance of power, war,—that was the old dreary cycle, now to be renewed unless we prevent it.
- ..... 8. A science tries to verify its hypotheses experimentally; therefore psychology is becoming increasingly scientific, since it is submitting its theories to laboratory tests.
- ..... 9. No A is B; some C is A; therefore some C is not B.
- ..... 10. That sign says: "Only ticket-holders will be admitted;" we can get in all right for we all have tickets.
- ..... 11. No A is B; some B is C; therefore no A is C.
- ..... 12. All organisms need food to live; this one will die for it is unable to assimilate food.

- .....13. All B is C; all A is B; therefore all A is C.
- .....14. During the war, all who were pro-German were against our government; therefore all who were against our government were pro-German.
- .....15. Only A is B; all B is C; therefore all A is C.
- .....16. Greece should have Constantinople and establish her capital where it was for a thousand years and more on the shores of the Bosphorus.
- .....17. Some A is not B; all A is C; therefore some C is not B.
- .....18. He must be a Christian for only Christians hold these views.
- .....19. Some things which are not A are B; no C is B; therefore some things which are not A are C.
- .....20. All M is P; no M is S; therefore no S is P.

TEST XI—ADDITION

Confidence	<u>4</u>	<u>602</u>	<u>7</u>	<u>603</u>	<u>5</u>	<u>900</u>	}	Row I
		↙ ↘		↙ ↘		↙ ↘		
		<u>76</u>	<u>97</u>	<u>73</u>	<u>69</u>	<u>789</u>	}	Row II
		<u>38</u>	<u>64</u>	<u>95</u>	<u>97</u>	<u>86</u>	}	
		↙ ↘		↙ ↘		↙ ↘		
		<u>469</u>	<u>736</u>	<u>694</u>	<u>365</u>	<u>921</u>	}	Row III
		<u>746</u>	<u>973</u>	<u>133</u>	<u>471</u>	<u>748</u>	}	
		↙ ↘		↙ ↘		↙ ↘		
		<u>8167</u>	<u>3584</u>	<u>4796</u>	}		Row IV	
		<u>9438</u>	<u>2831</u>	<u>5462</u>	}			
		<u>1781</u>	<u>9763</u>	<u>3697</u>	}			

## TEST XII—ETHICAL QUESTIONS

- ..... 1. Is capital punishment ever right?
- ..... 2. Should war be waged if it were certain that good would come of it?
- ..... 3. Is the Eighteenth Amendment an unjust restriction of personal freedom?
- ..... 4. Is it right for a schoolboy to report another whom he finds breaking a school regulation?
- ..... 5. Are business firms justified in forbidding the bobbing of hair among their women employees?
- ..... 6. Should details of crime be kept out of the newspapers?
- ..... 7. Should a college professor be allowed to teach doctrines which are subversive of the established order?
- ..... 8. Can this country rightly indulge in secret diplomacy?
- ..... 9. Is lynching ever justifiable?
- ..... 10. Should those found agitating an overthrow of the government of this country be allowed their liberty?
- ..... 11. Is it right to continue to maintain a high protective tariff?
- ..... 12. Is it ever right to lie?
- ..... 13. Should the Bible be taught in the public schools?
- ..... 14. Should the divorce laws be made less stringent?
- ..... 15. Is Oregon justified in passing a law to compel children to attend the public schools?
- ..... 16. Does the United States have a certain responsibility in the affairs of Europe?
- ..... 17. Should hazing be absolutely prohibited in boys' colleges?
- ..... 18. Should the censorship of motion pictures be entrusted to the good judgment of the public instead of to censors?
- ..... 19. If the trolley conductor neglects to take your fare, is it wrong to ride without paying?
- ..... 20. Can generosity be a fault?

## TEST XIII—CAUSAL JUDGMENTS

1. *Why are there so many traffic accidents every year? Because*
  - .....a. People are careless.
  - .....b. Traffic laws are not severe enough.
  - .....c. Licensing of drivers is not sufficiently restricted.
  - .....d. The war bred an indifference to human life.
2. *Why is this country in such a state of economic unrest? Because*
  - .....a. It lacks right leadership.
  - .....b. The labor unions are so strong and active.
  - .....c. Bolshevik propaganda is at work.
  - .....d. It has not entered the League of Nations.

3. *Why is not the advice of Steinmetz followed in the matter of using water power for electricity in this state, instead of burning coal for steam? Because*

- .....a. It would destroy the scenic beauty of the waterfalls.
- .....b. The financial interests prevent it.
- .....c. There is too much public apathy.
- .....d. The scheme wouldn't work.

4. *The present religious controversies are due to*

- .....a. The influences of Bryan's speeches on Evolution.
- .....b. A nation-wide religious revival.
- .....c. The utterances of such preachers as Grant and Fosdick.
- .....d. A scientific spirit which criticizes all dogma.

5. *Attempts at spiritistic communication have been so unsuccessful for*

- .....a. It is sacrilege to pry into the affairs of the Other World.
- .....b. The best scientific minds have not been at work at the problem.
- .....c. There is no spirit world.
- .....d. The gullibility of the public has made fraud more profitable than research.

6. *The course of history has been what it has because of*

- .....a. The influence of great men.
- .....b. The influence of the common people.
- .....c. The working out of economic laws.
- .....d. The political activities of the countries concerned.

7. *The present extravagance in this country is due to*

- .....a. A continuance of the post war reaction to thrift.
- .....b. The development of the psychology of advertising.
- .....c. The high wages everywhere being paid.
- .....d. The weakening of the moral and artistic fibre of the people.

8. *Hylan was re-elected Mayor because*

- .....a. He had been a good Mayor.
- .....b. He is skillful in the use of patronage.
- .....c. He is a Roman Catholic.
- .....d. He is allied with the money interests.

9. *The French went into Germany because*

- .....a. They thought they could thus collect their reparations.
- .....b. They intended to get a permanent foothold on German soil.
- .....c. They wished to control the coal supply of Europe.
- .....d. Poincare believed that only thus could he keep in power.

10. *The motion picture theatres are so frequented because*

- .....a. The entertainments are within the financial reach of all.
- .....b. Light and motion make a native appeal to the eye.
- .....c. The darkness affords opportunities for the amorous.
- .....d. The excitement and romance are compensation for the drabness of every-day existence.

## APPENDIX B

*Bibliography*

Such works as are mentioned in Chapter I are included here, as well as some others which in some cases are briefly annotated.

- Bagehot, W.—The Emotion of Conviction. *Literary Studies*, I, 412-17.
- Bain, A.—Emotions and Will. XI.  
Mental and Moral Science, Bk. iv, VIII.
- Brentano, F.—Psychologie vom Empirischen Standpunkte. II, 7.
- Bridges, J. W.—An Experimental Study of Decision Types and Their Mental Correlates. *Psychol. Monog.*, 1914, 17, 1-72.  
Parts of this experimentation relate to confidence.
- Cattell, J. M.—American Men of Science. New York: The Science Press, 1910.  
The Preface contains a description of the order-of-merit method of rating used, and references.
- Cooley, C. H.—Human Nature and the Social Order. Scribners, 1922. IX.
- Downey, J.—The Will-Profile. Bulletin No. 3, University of Wyoming, 1919.  
The Will-Temperament and Its Testing. World Book Co.  
Testing the Will-Temperament Tests. *School and Society*, August 5, 1922, 397, p. 161.  
A summary of investigations made to determine the validity of the tests.
- Filter, R. O.—An Experimental Study of Character Traits.  
*J. of Appl. Psychol.*, 1921, 5, 297-317.
- Fullerton and Cattell.—On the Perception of Small Differences.  
Publications of the University of Pennsylvania, 1892, 2.
- Garrett, H. E.—A Study of the Relation of Accuracy to Speed. *Arch. of Psychol.*, 1922, 56.  
Incidental conclusions on confidence.
- Griffing, H.—On Sensations from Pressure and Impact. *Psychol. Monog.*, 1895, 1.
- Henmon, V. A. C.—The Relation of the Time of a Judgment to Its Accuracy. *Psychol. Rev.*, 1911, 18, 186-201.
- Hollingworth, H. L.—Experimental Studies in Judgment. *Arch. of Psychol.*, 1913, 29.  
Judging Human Character. D. Appleton, 1922. IV, IX, XII.
- Hume, D.—Treatise on Human Nature, Bk. I, Part iii, Secs. 7-10.
- James, W.—Psychology. Vol. ii. Ch. XXI.
- Janet, P.—Les Obsessions et la Psychasthénie. III.
- Jastrow, J.—The Psychology of Conviction. Houghton Mifflin, 1918.
- Lindsay, J.—Psychology of Belief, *Arch. für Syst. Phil.*, 1910, 16, 293-309.  
Contains a careful resumé of the views of different psychologists on belief from the time of Hume up to 1910.
- McDougall, W.—Belief as a Derived Emotion. *Psychol. Rev.*, 1921, 28, 315-328.  
Outline of Psychology. Scribners, 1923. XII, XIV.
- Martin, A. H.—An Empirical Study of the Factors and Types of Voluntary Choice. *Arch. of Psychol.*, 1922, 51.  
An introspective inquiry using the method of reversal.
- Metcalf, J. T.—An Experimental Study of the Conscious Attitudes of Certainty and Uncertainty. *Psychol. Monog.*, 1917, 23, 181-240.
- Mill, J. S.—Analysis of Phenomena of Human Mind. I. Ch. XI.  
See note, p. 402.



- Moore, H. T.—The Comparative Influence of Majority and Expert Opinion. *Am. J. of Psychol.*, 1921, 32, 16-20.  
The method of reversal is employed with 6 subjects in 3 kinds of situations.
- Moore, H. T. and Gilliland, A. R.—The Measurement of Aggressiveness. *J. of Appl. Psychol.*, 1921, 5, 97-118.
- Okabe, T.—An Experimental Study of Belief. *Am. J. of Psychol.*, 1910, 21, 563-596.  
An introspective study, containing also a brief review of psychological writings on belief.
- Pillsbury, W. B.—*Psychology of Reasoning*. D. Appleton, 1910. II.
- Roback, A. A.—*Behaviorism and Psychology*. Cambridge: University Book Store, XII.  
The Psychology of Belief. *Psychol. Bull.*, 1920, 17, 53-54.
- Rogers, A. K.—What is Truth? Yale Univ. Press, 1923. 1-54.
- Rousmaniere, F. H.—Certainty and Attention. *Harvard Psychol. Studies*, 1906, 2, 277-291.  
An introspective study of the nature of the feeling of certainty.
- Ruger, H. A.—The Psychology of Efficiency. *Arch. of Psychol.*  
Includes the influence of attitudes in problem solving.
- Russel, B.—The Analysis of Mind.
- Stern, W.—Abstracts of Lectures on the Psychology of Testimony, *Am. J. of Psychol.*, 1910, 21, 270-275.  
Discusses the relation of confidence to accuracy.
- Stout, G. F.—*Analytic Psychology*. London: George Allen & Unwin, 97-111.
- Strong, E. K., Jr.—The Effect of Length of Series upon Recognition Memory. *Psychol. Rev.*, 1912, 19, 447-462.
- Sully, J.—Sensation and Intuition, Essay IV.
- Sumner, F. B.—A Statistical Study of Belief. *Psychol. Rev.*, 1898, 5, 616-631.
- Titchener, E. B.—*Experimental Psychology of the Higher Thought Processes*. Macmillan, 1909, 250-258.
- Watson, J. B.—Is Thinking Merely the Action of Language Mechanisms? *Brit. J. of Psychol.*, 1920, 11, 87-104.
- Wells, W. R.—The Biological Foundations of Belief. Badger, 1921.
- Williamson, G. F.—Individual Differences in Belief, Measured and Expressed by Degrees of Confidence. *J. of Philos. Psychol.*, etc., 1915, 12, 127-137.
- Woodworth & Poffenberger—*Experimental Psychology*. XXI (in preparation).



BF 21.A7 no67



3 9358 00040924 0



3 9358 0004924 0

## DIRECTORY OF AMERICAN PSYCHOLOGICAL PERIODICALS

- American Journal of Psychology**—Ithaca, N. Y.: Morrill Hall.  
Subscription \$6.50. 600 pages annually. Edited by E. B. Titchener.  
Quarterly. General and experimental psychology. Founded 1887.
- Pedagogical Seminary**—Worcester, Mass.: 950 Main Street.  
Subscription \$5. 575 pages annually. Edited by G. Stanley Hall.  
Quarterly. Pedagogy and educational psychology. Founded 1891.
- Psychological Review**—Princeton, N. J.: Psychological Review Company.  
Subscription \$4.25. 480 pages annually.  
Bi-monthly. General. Founded 1894. Edited by Howard C. Warren.
- Psychological Bulletin**—Princeton, N. J.: Psychological Review Company.  
Subscription \$5. 720 pages annually. Psychological literature.  
Monthly. Founded 1904. Edited by Shepherd I. Franz.
- Psychological Monographs**—Princeton, N. J.: Psychological Review Company.  
Subscription \$5.50 per vol., 500 pp. Founded 1895. Ed. by James R. Angell.  
Published without fixed dates, each issue one or more researches.
- Psychological Index**—Princeton, N. J.: Psychological Review Company.  
Subscription \$1.50. 200 pp. Founded 1895. Edited by Madison Bentley.  
An annual bibliography of psychological literature.
- Journal of Philosophy**—New York: Sub-Station 84.  
Subscription \$4., 728 pages per volume. Founded 1904.  
Bi-weekly. Edited by F. J. E. Woodbridge and Wendell T. Bush.
- Archives of Psychology**—Sub-Station 84, New York: Archives of Psychology.  
Subscription \$5. 500 pp. annually. Founded 1906. Edited by R. S. Woodworth.  
Published without fixed dates, each number a single experimental study.
- Journal of Abnormal Psychology and Social Psychology**—Boston.  
Subscription \$5. Richard G. Badger. Edited by Morton Prince.  
Bi-Monthly. 432 pages annually. Founded 1906. Abnormal and social.
- Psychological Clinic**—Philadelphia: Psychological Clinic Press.  
Subscription \$2.50. 288 pages. Edited by Lightner Witmer. Founded 1907.  
Without fixed dates (9 numbers). Orthogenics, psychology, hygiene.
- Training School Bulletin**—Vineland, N. J.: The Training School.  
Subscription \$1. 160 pages annually. Edited by E. R. Johnstone. Founded 1904.  
Monthly (10 numbers). Psychology and training of defectives.
- Journal of Educational Psychology**—Baltimore: Warwick & York.  
Subscription \$4. 576 pages annually. Founded 1910.  
Monthly (9 numbers). Managing Editor, Harold O. Rugg.  
(Educational Psychology Monographs.  
Published separately at varying prices. Same publishers.)
- Comparative Psychology Monographs**—Baltimore: Williams & Wilkins Co.  
Subscription \$5. 500 pages per volume. Edited by W. S. Hunter.  
Published without fixed dates, each number a single research.
- Psychoanalytic Review**—Washington, D. C.: 3617 10th Street, N. W.  
Subscription \$6. 500 pages annually. Psychoanalysis.  
Quarterly. Founded 1913. Edited by W. A. White and S. E. Jelliffe.
- Journal of Experimental Psychology**—Princeton, N. J.  
Psychological Review Company. 480 pages annually. Experimental.  
Subscription \$4.25. Founded 1916. Bi-monthly. Edited by John B. Watson.
- Journal of Applied Psychology**—Worcester, Mass.: Florence Chandler.  
Subscription \$4. 400 pages annually. Founded 1917.  
Quarterly. Edited by James P. Porter and William F. Book.
- Journal of Comparative Psychology**—Baltimore: Williams & Wilkins Co.  
Subscription \$5. 500 pages per volume. Founded 1921.  
Bi-monthly. Edited by Knight Dunlap and Robert M. Yerkes.